1981

Bounded Stopping Times for a Class of Sequential Bayes Tests

R. H. Berk
Rutgers University

Lawrence D. Brown
University of Pennsylvania

Arthur Cohen
Rutgers University

Recommended Citation

Berk, R. H., Brown, L. D., & Cohen, A. (1981). Bounded Stopping Times for a Class of Sequential Bayes Tests. *The Annals of Statistics, 9* (4), 834-845. http://dx.doi.org/10.1214/aos/1176345523

Bounded Stopping Times for a Class of Sequential Bayes Tests

Abstract
Consider the problem of sequentially testing a null hypothesis vs an alternative hypothesis when the risk function is a linear combination of probability of error in the terminal decision and expected sample size (i.e., constant cost per observation.) Assume that the parameter space is the union of null and alternative, the parameter space is convex, the intersection of null and alternative is empty, and the common boundary of the closures of null and alternative is nonempty and compact. Assume further that observations are drawn from a *p*-dimensional exponential family with an open *p*-dimensional parameter space. Sufficient conditions for Bayes tests to have bounded stopping times are given.

Keywords
sequential tests, hypothesis testing, Bayes test, exponential family, stopping times, monotone likelihood ratio

Disciplines
Statistics and Probability

The Annals of Statistics
1981, Vol. 9, No. 4, 834–845

BOUNDED STOPPING TIMES FOR A CLASS OF SEQUENTIAL BAYES TESTS

BY R. H. BERK[1], L. D. BROWN[2], AND ARTHUR COHEN[2]

Rutgers University, Cornell University, and Rutgers University

Consider the problem of sequentially testing a null hypothesis vs an alternative hypothesis when the risk function is a linear combination of probability of error in the terminal decision and expected sample size (i.e., constant cost per observation.) Assume that the parameter space is the union of null and alternative, the parameter space is convex, the intersection of null and alternative is empty, and the common boundary of the closures of null and alternative is nonempty and compact. Assume further that observations are drawn from a p-dimensional exponential family with an open p-dimensional parameter space. Sufficient conditions for Bayes tests to have bounded stopping times are given.

1. Introduction and summary. Consider the problem of sequentially testing a null hypothesis vs an alternative hypothesis when the risk function is a linear combination of probability of error in the terminal decision and expected sample size (i.e., constant cost for each observation). Assume that the parameter space is the union of null and alternative, the parameter space is convex, the intersection of null and alternative is the empty set, and the common boundary of the closures of null and alternative is nonempty and compact. Assume further that observations are drawn from a p-dimensional exponential family with an open p-dimensional parameter space. Sufficient conditions for Bayes tests to have bounded stopping times are given. The main conditions are as follows: For $p = 1$, the supports of the prior probability measures, conditioned on the null and alternative spaces, both contain the common boundary. For $p > 1$, the supports of the prior probability measures, conditioned on the null and alternative spaces, both contain some neighborhood of the common boundary intersected with their respective spaces. Extensions in various directions are given. Most importantly, the results hold for a variety of loss functions.

Knowledge that a Bayes procedure has a bounded stopping time is important. In addition to being a desirable property of a sequential test, it is needed to compute a Bayes procedure explicitly. Only in such cases can the backward recursion method be used exactly. Another advantage is that bounded stopping time implies, for many models, that the risk functions (under null and alternative) of Bayes tests are continuous. The fundamental lemmas developed in Section 2 will be used in a subsequent paper to establish continuity of risks for Bayes tests in a slightly different setting. See Berk, Brown, and Cohen (1981). This latter result will relate and add to some complete class results in Brown, Cohen, and Strawderman (1980). The main result of this paper will also be applied to demonstrate that certain weight function sequential tests are inadmissible. See Brown and Cohen (1981). The results of the paper are of great importance in increasing our understanding of sequential testing.

Received April, 1979; revised August, 1980.

[1] Research supported by N.S.F. Grant No. MCS-76-82618-A01

[2] Research supported by N.S.F. Grant No. MCS-75-23343-A02

AMS 1970 *subject classifications.* Primary 62L10; secondary 62C10, 52L15.

Key words and phrases. Sequential tests, hypothesis testing, Bayes test, exponential family, stopping times, monotone likelihood ratio.

There are some previous discussions in the literature of sequential hypothesis testing where it has been possible to conclude that certain Bayes tests have a bounded stopping time. Sobel (1952) and Schwarz (1962) study priors whose support is the entire one dimensional parameter space. However in their formulations there exists an indifference zone. Parameters in the indifference zone are part of the model even though the loss due to any terminal decision is zero if the parameter is in the indifference zone. In our model there is no indifference zone. The property of a bounded stopping time for a Bayes test is relatively easy to prove when there is an indifference zone. Roughly the argument is as follows: Suppose the support of the prior is the entire parameter space. Then in the setting of observations drawn from a one dimensional exponential family it can be shown that the posterior distribution, given the sample through stage n, will become highly concentrated as $n \to \infty$, regardless of the actual sample value observed. There are three possibilities. The posterior is highly concentrated in the null space, in the alternative space, or near the boundary of the two. In the first two cases a wrong terminal decision has low a posteriori probability and hence the contribution to the posterior risk due to making a wrong terminal decision is less than the cost of another observation. Hence the Bayes procedure stops. In the third case the posterior probability of making a wrong terminal decision may be high, but because of the loss structure (indifference zone with zero loss), the posterior risk is small. It thus risks less to stop and make any terminal decision, than it costs to continue sampling.

With constant loss for a wrong terminal decision, which we assume, the preceding argument fails in the third case. A different type of argument is required. The argument for this case is based on Lemmas 2.1 and 2.2. These lemmas involve looking many steps ahead, rather than just one as in Sobel (1952) and Schwarz (1962). These lemmas show that in this third case even if sampling were to continue for a long time the situation would not be likely to improve much. With high probability, the new posterior will be nearly as uninformative as the old one, concerning whether the null or alternative is true. Hence the posterior risk due to wrong terminal decision will still (with high probability) remain almost as large. It pays to stop immediately rather than sample for such a long time for a relatively small expected reduction in terminal risk.

The question of bounded stopping times for Bayes tests with our loss structure has been studied by Ray (1965). He studies some continuous and discrete one dimensional exponential family distributions. He gives a sufficient condition in terms of the prior distribution for a Bayes test to have a bounded stopping time. His most usable condition is for one sided tests. However, his condition usually is extremely difficult to verify. Ray also develops formulas for explicit bounds on the stopping time. The formulas are difficult to use except in some very special cases. The conditions here apply to a very large class of priors and require no other verification.

Extensions in several directions are possible. First the exponential family assumption in the one dimensional case can be replaced by requiring that the observations are independent and identically distributed and that the Bayes tests are based on a sequence of sufficient statistics whose distributions is monotone likelihood ratio. This then would yield a result for the uniform distribution with unknown scale parameter. (See the end of Section 3.)

The requirement that the parameter space is convex can be altered. (See Remark 3.1.) The results can be developed for more general loss functions. Generalizations of the 0-1 loss due to terminal decision and generalization of the constant cost of observations are discussed at the end of Section 3. For the one dimensional model, the assumption that the intersection of null and alternative be the empty set can be relaxed provided the condition on the prior is the one given for $p > 1$. (See Remark 3.2.) Also a result for testing the noncentrality parameter of a chi-square random variable is given in Remark 3.3.

Definitions and the fundamental lemmas are in Section 2. The theorem is given in Section 3 and extensions are given at the end of Section 3.

2. Definitions, Preliminaries, and Main Lemmas. The elements of the problem are as follows: Θ is the parameter space with typical element θ. The null space is $\Theta_1 \subset \Theta$ and the alternative space is $\Theta_2 \subset \Theta$. Assume Θ, Θ_1, Θ_2 are measurable subsets of Euclidean space R^p. In this paper we assume $\Theta = \Theta_1 \cup \Theta_2$ and $\Theta_1 \cap \Theta_2 = \phi$, the null set. (See Remark 3.2 where $\Theta_1 \cap \Theta_2$ need not be a null set.) The closures of Θ_1, Θ_2, and Θ are denoted respectively by $\bar{\Theta}_1$, $\bar{\Theta}_2$, and $\bar{\Theta}$. The action space $\mathscr{A}$ consists of pairs (n, τ) where n is the value of the stopping time, N, and τ is 1 or 2, depending on whether Θ_1 is accepted or rejected. The loss function, denoted by $L(\theta, (n, \tau)) = cn$ if $\theta \in \Theta_\tau$ and $L(\theta, (n, \tau)) = cn + 1$ if $\theta \notin \Theta_\tau$. Here $c > 0$ represents the cost of each individual observation.

The observation available to the statistician at stage j is an observation on a random variable X_j whose realization is denoted by x_j. We let $\mathbf{x} = (x_1, x_2, \cdots)$ denote the sequence of realized observations and $\mathbf{x}_j = (x_1, x_2, \cdots, x_j)$. The corresponding random variables (measurable mappings) are denoted by $\mathbf{X}$ and $\mathbf{X}_j$ respectively. The sample space of $\mathbf{X}_j$ is denoted by $\mathscr{X}_j$. Throughout, the affix j is omitted when $j = 1$. We assume that there is a σ-finite measure μ which dominates the family $\{P_\theta(\cdot), \theta \in \Theta\}$ of probability measures for $\mathbf{X}$ in the following sense: For each $j = 1, 2, \cdots$, over the σ-field generated by $\mathbf{X}_j$, the measure P_θ is dominated by μ. Write $f_{\theta j} = dP_\theta/d\mu$ relative to this σ-field. When $j = 1$, write $f_{\theta 1} = f_\theta$. When Θ is not closed we will assume that the family, $\{P_\theta, \theta \in \Theta\}$, can be extended to a family, $\{P_\theta, \theta \in (\Theta \cup \Omega) = \Lambda\}$, where $\Omega = \bar{\Theta}_1 \cap \bar{\Theta}_2$, and that the families of densities $f_{\theta j}$, $\theta \in \Lambda$, exist and have certain properties (essentially continuity properties) to be specified later. We assume the observations are independent and identically distributed.

A prior probability measure on Θ, denoted by $\Gamma(\cdot)$, will be represented by a mixture expressed as $\pi_1\Gamma_1(\cdot) + \pi_2\Gamma_2(\cdot)$. Here, if T is a random variable with distribution Γ, then π_1 is the probability that $T \in \Theta_1$ and Γ_1 represents the conditional distribution of T, given $T \in \Theta_1$; similarly for Γ_2. A prior may be represented as $(\pi_1, \Gamma_1(\cdot), \Gamma_2(\cdot))$.

A decision function δ may be expressed as a set of nonnegative functions $\delta_{ij}(\mathbf{x}_j) \geq 0$, $(i = 0, 1, 2; j = 1, 2, \cdots)$ defined for all $\mathbf{x}_j \in \mathscr{X}_j$ such that $\sum_{i=0}^{2} \delta_{ij}(\mathbf{x}_j) = 1$. The quantities $\delta_{ij}(\mathbf{x}_j)$, $i = 0, 1, 2$, represent respectively, the probability of taking another observation, accepting H_1, and accepting H_2 when j observations have been taken and $\mathbf{X}_j = \mathbf{x}_j$ is observed.

The risk function is $E_\theta L(\theta, \delta)$ and the expected risk is $EL(\theta, \delta)$. A Bayes procedure minimizes $EL(\theta, \delta)$. For our study a Bayes procedure always exists; see Le Cam (1955). The Bayes procedure with respect to the prior Γ will be denoted by δ_Γ and its stopping time, by N_Γ.

Now recall $\Omega = \bar{\Theta}_1 \cap \bar{\Theta}_2$. For the problems treated here we will assume that Ω, the common boundary of the null and alternative spaces is nonempty, compact, and has typical element ω. Let $M(\epsilon, \theta)$ be an ϵ-neighborhood of θ, defined as the set of all $t \in \Theta$ such that $\|\theta - t\| < \epsilon$, for any $\theta \in \Theta$. ($\|\cdot\|$ is Euclidean norm). For any set $A \subset \Theta$, let $M(\epsilon, A) = \cup_{\theta \in A} M(\epsilon, \theta)$ be the ϵ-neighborhood of the set A.

Define the support of a probability measure $\Gamma(\cdot)$, as is usual, by supp $\Gamma = \cap \{C : \Gamma(C) = 1, C \text{ closed}\}$. Further for $i = 1, 2$, let

$$g_{in}(\mathbf{x}_n) = \int_{\Theta_i} f_{\theta n}(\mathbf{x}_n)\Gamma_i(d\theta) \Big/ \left[\pi_1 \int_{\Theta_1} f_{\theta n}(\mathbf{x}_n)\Gamma_1(d\theta) + \pi_2 \int_{\Theta_2} f_{\theta n}(\mathbf{x}_n)\Gamma_2(d\theta)\right] \tag{2.1}$$

so that $\pi_i g_{in}(\mathbf{x}_n)$ represents the conditional probability that $\theta \in \Theta_i$, given $\mathbf{X}_n = \mathbf{x}_n$. Finally let the measure P be the mixture of P_θ when Γ is the mixing distribution. We are now prepared to prove the main lemmas. Note in the statements of these lemmas that the requirements on the densities are minimal and even these requirements are not always needed. For ease of presentation we omit some subscripts when no confusion will arise.

LEMMA 2.1. *Suppose $X_1, X_2, \cdots$ are independent and identically distributed. Assume that for almost every fixed $x \in \mathscr{X}$, the density $f_\theta(x)$ is continuous in θ at $\theta = \omega$ for each $\omega \in \Omega$. Assume also that $\sup_{\theta \in \Lambda} f_\theta(x) < \infty$ for almost all $x \in \mathscr{X}$. Then for every positive constant $0 < \xi < 1$, there exists an $\epsilon > 0$ such that, if for some $\omega \in \Omega$,*

$$\Gamma_i(M(\epsilon,\omega)\cap\Theta_i) > 1-\epsilon, \qquad i=1,2, \tag{2.2}$$

then

$$P_\theta(\pi_i(1-\xi) < \pi_i g_{i1}(X_1),\ i=1,2) > 1-\xi, \qquad \theta\in M(\epsilon,\omega). \tag{2.3}$$

Hence for the given ξ and any positive integer k, there is an $\epsilon_k > 0$ so that if (2.2) *holds with $\epsilon = \epsilon_k$,*

$$P(\pi_i(1-\xi) > \pi_i g_{ij}(\mathbf{X}_j),\ i=1,2, \quad \textit{for every} \quad j\le k) \ge (1-k\xi)(1-\epsilon_k). \tag{2.4}$$

PROOF. Due to the continuity and boundedness assumptions, for each $\omega\in\Omega$ there is an $\epsilon(\omega)>0$, a set $E(\omega)\subset\mathscr{X}$, and a $\delta(\omega)>0$ such that the following conditions are all satisfied: $P_\omega(E(\omega)) > 1-3\xi/4$; for all $x\in E(\omega)$, $f_\omega(x) > \delta(\omega)$ and $\sup_{\theta\in\Lambda} f_\theta(x) < 1/\delta(\omega)$; $\inf\{[f_\theta(x)/f_\omega(x)];\ x\in E(\omega),\ \theta\in M(\epsilon(\omega),\omega)\cap\Lambda\} > 1-\xi/4$; and $\sup\{[f_\theta(x)/f_\omega(x)];\ x\in E(\omega),\ \theta\in M(\epsilon(\omega),\omega)\cap\Lambda\} < 1+\xi/4$. It follows that for $\theta\in M(\epsilon(\omega),\omega)$

$$\begin{aligned} P_\theta(E(\omega)) &= \int_{E(\omega)} f_\theta(x)\mu\,(dx) \\ &\ge (1-\xi/4)\int_{E(\omega)} f_\omega(x)\mu\,(dx) \\ &= (1-\xi/4)P_\omega(E(\omega)) > (1-\xi). \end{aligned} \tag{2.5}$$

Also, letting $\Lambda_i = \Theta_i \cup \Omega$,

$$\begin{aligned} &\inf\{f_\lambda(x)/f_\theta(x):\lambda\in M(\epsilon(\omega),\omega)\cap\Lambda_1,\ \theta\in M(\epsilon(\omega),\omega)\cap\Lambda_2\} \\ &\qquad > (1-\xi/4)/(1+\xi/4) > 1-\xi/2, \end{aligned} \tag{2.6}$$

for every $x\in E(\omega)$. Note,

$$\begin{aligned} K(\omega) &= \sup\{f_\lambda(x)/f_\theta(x):\lambda\in\Lambda_2,\ \theta\in M(\epsilon(\omega),\omega)\cap\Lambda_2,\ x\in E(\omega)\} \\ &\le [1/\delta(\omega)]/[\delta(\omega)(1-\xi/4)] < \infty. \end{aligned} \tag{2.7}$$

Now, $\Omega\subset\cup_{\omega\in\Omega} M(\epsilon(\omega)/2,\omega)$. Since Ω is compact, there is a finite subcover, say $\{M(\epsilon(\omega_j)/2,\omega_j),\ j=1,2,\cdots,J\}$ such that $\Omega\subset\cup_{j=1}^{J} M(\epsilon(\omega_j)/2,\omega_j)$. Choose $\epsilon'>0$ and $\epsilon' < \min_{1<j\le J}\{\epsilon(\omega_j)/2\}$. Then for every $\omega\in\Omega$, there is an ω_j such that $M(\epsilon',\omega)\subset M(\epsilon(\omega_j),\omega_j)$. Let $K=\max_{1\le j\le J}K(\omega_j)$ and choose ϵ such that $0<\epsilon<\epsilon'$, $(1-\xi/2)(1-\epsilon) > (1-2\xi/3)$, $K\epsilon/(1-\epsilon) < \xi/3$.

Suppose $\Gamma_i(M(\epsilon,\omega)\cap\Theta_i) > 1-\epsilon$, $i=1,2$. Then for some j, $M(\epsilon,\omega)\subset M(\epsilon(\omega_j),\omega_j) = M_j$ and $\Gamma_i(M_j\cap\Theta_i) > 1-\epsilon$. It then follows from (2.1), (2.6), and (2.7) that for every $x\in E = E(\omega_j)$,

$$\begin{aligned} [g_{11}(x)/g_{21}(x)] &= \int_{\Theta_1} f_\theta(x)\Gamma_1(d\theta)\Big/\int_{\Theta_2} f_\theta(x)\Gamma_2(d\theta) \\ &\ge \int_{M_j\cap\Theta_1} f_\theta(x)\Gamma_1(d\theta)\Big/\left[\int_{M_j\cap\Theta_2} f_\theta(x)\Gamma_2(d\theta)\left(1+\frac{\int_{\Theta_2-M_j} f_\theta(x)\Gamma_2(d\theta)}{\int_{M_j\cap\Theta_2} f_\theta(x)\Gamma_2(d\theta)}\right)\right] \\ &\ge (1-\xi/2)(1-\epsilon)/[1+K\epsilon/(1-\epsilon)] \\ &\ge (1-2\xi/3)/(1+\xi/3) \ge (1-\xi). \end{aligned} \tag{2.8}$$

The $\epsilon(\omega)$ determined to satisfy (2.6) and (2.7) can be chosen to satisfy (2.6) and (2.7) with the indices 1 and 2 interchanged. Hence interchanging the indices 1 and 2 in (2.8) implies also that $g_{21}(x)/g_{11}(x) \ge (1 - \xi)$ for every $x \in E = E(\omega_j)$. Using the fact that $\pi_1 g_{11}(x) + \pi_2 g_{21}(x) = 1$, it is easily seen that (2.8), and (2.8) with indices 1 and 2 interchanged, imply (2.3). We may repeat the above argument with $f_{\theta k}$ replacing f_θ and conclude that for each $k \ge 1$, there is an ϵ_k so that (2.2) implies (2.3), provided ϵ is replaced by ϵ_k in both expressions. Without loss of generality, we may assume $\epsilon_1 \ge \epsilon_2 \ge \cdots \ge \epsilon_k$. Since P is the mixture of P_θ when Γ is the mixing distribution, (2.2) holding with $\epsilon = \epsilon_k$, implies (2.4) holds as well by applying one of De Morgan's Laws.

REMARK 2.1. Equation (2.4) of Lemma 2.1 can still be proved if the assumptions on $f_\theta(x)$ are weakened to the requirement that $f_\theta(X_1) \to f_\omega(X_1)$ in P_ω-probability as $\theta \to \omega$, for each $\omega \in \Omega$. Furthermore independence of $X_1, X_2 \cdots$ can be relaxed provided the continuity and boundedness assumptions hold for $f_{\theta j}$. Independence is needed in the succeeding lemma where Lemma 2.1 is applied.

REMARK 2.2. In Lemma 2.1 the integer k should be thought of as a number of prospective future observations. The lemma says that given k and ξ, there exists an ϵ_k such that if a prior concentrates most of its mass around some boundary point $\omega \in \Omega$ (i.e., $\Gamma_i(M(\epsilon_k, \omega) \cap \Theta_i) > 1 - \epsilon_k$) then the random variables measuring the posterior probabilities of lying in Θ_i in the future (anytime during the next k observations) will both be close to their respective prior probabilities. In other words, if the prior concentrates heavily near the boundary, then the data will not easily be able to discriminate between null and alternative at anytime during the next k steps. This fact will be used in the next lemma. There the idea is that if the posterior distributions concentrate their mass around a point in the boundary, then improvement in the posterior loss due to terminal decision for the next k future observations is expected to be so small that it is better to stop now rather than incur additional loss due to cost of these future observations.

Now let $\tau_n(\mathbf{x}_n) = \min_{i=1,2}(\pi_i g_{in}(\mathbf{x}_n))$, let $U_n(\mathbf{x}_n)$ denote the minimum conditional risk given sampling stops at stage n after $\mathbf{x}_n$ has been observed. That is $U_n(\mathbf{x}_n) = cn + \tau_n(\mathbf{x}_n)$. Also let $V_n(\mathbf{x}_n)$ denote the minimal conditional expected risk given $\mathbf{x}_n$ and that sampling continues at least to stage $(n + 1)$. ($V_n(\mathbf{x}_n)$ can be expressed mathematically in a recursive way. (See Brown, Cohen, and Strawderman (1980), pages 383–385 where V_n is called $\beta_{n,k}$.) Further, let $\Gamma_i(\cdot \mid \mathbf{x}_n)$, $i = 1, 2$ denote the conditional (posterior) probability measures on Θ_i given $\mathbf{x}_n$ and that $\theta \in \Theta_i$.

LEMMA 2.2. *Assume the conditions of Lemma* 2.1 *hold. Let* $\Gamma = \pi_1\Gamma_1 + \pi_2\Gamma_2$ *be a prior probability measure. Let* $k \ge 1/2c$, $0 < \xi < 2c/(k + 1)$, *and* $\epsilon > 0$ *be smaller than* ϵ_k *in Lemma* 2.1 *and also small enough so that*

$$\epsilon < 2c - (k + 1)\xi. \tag{2.9}$$

Suppose $\mathbf{X}_n = \mathbf{x}_n$ *has been observed and for some* $\omega \in \Omega$

$$\Gamma_i(M(\epsilon, \omega) \cap \Theta_1 \mid \mathbf{X}_n = \mathbf{x}_n) > 1 - \epsilon, \qquad i = 1, 2. \tag{2.10}$$

Then δ_Γ *stops at time n.*

PROOF. First note that the independence of $X_1, X_2, \cdots$ implies that a restatement of (2.4) is

$$\begin{aligned} &P(\pi_i g_{in}(\mathbf{x}_n)(1 - \xi) < \pi_i g_{ij}(\mathbf{X}_j),\ i = 1, 2; \\ &\qquad\qquad (n + 1) \le j \le (n + k) \mid \mathbf{X}_n = \mathbf{x}_n) \ge (1 - k\xi)(1 - \epsilon). \end{aligned} \tag{2.11}$$

The definition of $\tau_n(\mathbf{x}_n)$ and (2.11) in turn imply

$$P(\tau_j(\mathbf{X}_j) > (1 - \xi)\tau_n(\mathbf{x}_n);\ (n + 1) \le j \le (n + k) \mid \mathbf{X}_n = \mathbf{x}_n) \ge (1 - k\xi)(1 - \epsilon). \tag{2.12}$$

Now assume $\mathbf{x}_n$ has been observed. Let N' denote the stopping time of the best procedure which continues at least to stage $(n + 1)$ after observing $\mathbf{x}_n$. Let $\lambda = P(N' \geq n + k + 1 \mid \mathbf{X}_n = \mathbf{x}_n)$. Then

$$
\begin{aligned}
V_n(\mathbf{x}_n) &\geq cn + c + ck\lambda + E(\tau_{N'}(\mathbf{X}_{N'}) \cdot 1_{(N'<n+k+1)} \mid \mathbf{X}_n = \mathbf{x}_n) \\
&\geq cn + c + (1 - k\xi)(1 - \epsilon)\tau_n(\mathbf{x}_n)(1 - \xi)(1 - \lambda) + ck\lambda,
\end{aligned}
\tag{2.13}
$$

by virtue of (2.12). Recall that $ck \geq \frac{1}{2} > \tau_n(\mathbf{x}_n)(1 - \xi)$. Hence the right hand side of (2.13) is minimized as a function of $\lambda \geq 0$ by the choice $\lambda = 0$, and

$$
\begin{aligned}
V_n(\mathbf{x}_n) &\geq cn + c + (1 - k\xi)(1 - \epsilon)(1 - \xi)\tau_n(\mathbf{x}_n) \\
&\geq cn + c + (1 - (k + 1)\xi - \epsilon)\tau_n(\mathbf{x}_n) \\
&> cn + \tau_n(\mathbf{x}_n) = U_n(\mathbf{x}_n),
\end{aligned}
\tag{2.14}
$$

by (2.9) since $\tau_n \leq \frac{1}{2}$. The inequality in (2.14) implies the validity of the lemma. □

3. Exponential family. In this section we assume that the observations are independent and identically distributed from an exponential family. The random vector $X \in R^p$. Here the sequence $\{\bar{X}_n\}$, $\bar{X}_n = \sum_{j=1}^n X_j/n$, is sufficient, transitive, and for $\bar{X}_n$, its family of probability distributions $P_{\theta n}(x)$, gives rise to densities of the form

$$\bar{f}_{\theta n}(x) = e^{n(\theta' x - \psi(\theta))}, \tag{3.1}$$

with respect to some measure μ_n. Note that in this section x will represent the realization $\bar{x}_n$ of the sufficient statistic and will be used as the argument of all functions in Section 2 whose argument there was $\mathbf{x}_n$. The parameter space Θ is assumed to be a convex subset of the natural parameter space $\boldsymbol{\Theta} = \{\theta : e^{\psi(\theta)} = \int e^{\theta' x}\, d\mu_1(x) < \infty\}$. We assume $\boldsymbol{\Theta}$ is open. The expectation of $\bar{X}_n$,

$$\beta \equiv E_\theta \bar{X}_n = b(\theta) \tag{3.2}$$

exists finitely for all $\theta \in \Theta$. Let B denote the expectation parameter space. The mapping b of θ to β is a 1-1 continuous map. Hence we may reformulate the original problem of testing Θ_1 vs Θ_2 in terms of B_1 vs B_2 where B_1 and B_2 are the images of Θ_1 and Θ_2. We let D be the image of Ω; D is also compact. Note that a prior distribution defined on Θ determines uniquely a prior distribution on B. (In some cases we may prefer anyway, to test a hypothesis concerning β and so we may consider placing the prior distribution directly on B or even $\bar{B}$ if we wish.) We will utilize the representation of the density in (3.1) given in Efron (1978), page 366, namely

$$\bar{f}_{\beta n}(x) = \bar{f}_{xn}(x) e^{-n I(x, \beta)}, \tag{3.3}$$

where $\bar{f}_{xn}$ indicates the density of $\bar{X}_n$ when $\beta = x$, and

$$I(x, \beta) = (v(x) - v(\beta))' x - [\psi(v(x) - \psi(v(\beta))], \tag{3.4}$$

where v is b^{-1}. Note that $I(x, \beta) = (\theta_x - \theta)' x - [\psi(\theta_x) - \psi(\theta)] = \hat{I}(x, \theta)$ where $\theta_x = v(x)$, $\theta = v(\beta)$, and as such is equal to the Kullback-Leibler distance $I^*(\theta_x, \theta)$. (See Efron (1978), pages 364 and 366; our $I^*(\theta_x, \theta)$ corresponds to Efron's $I(\alpha_1, \alpha_2)$.) Furthermore, $I^*(\theta_x, \theta)$, defined as the right-hand side of (3.4) is a convex function of θ for fixed θ_x. We utilize a simple result of Kallenberg (1978) page 40, which states that if C is a compact subset of $\boldsymbol{\Theta}$, then there exist positive constants A_1, A_2, A_3, A_4 depending only on C, such that for every $\theta_x, \theta \in C (\theta_x \neq \theta)$

$$A_1 \leq \| x - \beta \| / \| \theta_x - \theta \| \leq A_2, \tag{3.6}$$

$$A_3 \leq I^*(\theta_x, \theta) / \| \theta_x - \theta \|^2 \leq A_4. \tag{3.7}$$

Note (3.6) and (3.7) imply that there exist positive constants A_5, A_6 so that

$$A_5 \le I(x, \beta)/\|x - \beta\|^2 \le A_6. \tag{3.8}$$

We choose a compact $C \subset \boldsymbol{\Theta}$ so that H, the image of C under the expectation map b, contains $M(\epsilon, D)$ where ϵ is chosen as in Lemma 2.2. Since D is compact there will be no difficulty in determining C and H for the proof in the theorem to follow.

It will be convenient in the theorem to follow to let Γ, the prior distribution, be defined directly on B instead of on Θ. In fact let $\boldsymbol{\Gamma}$ denote the prior measure on Θ induced by Γ. We list two conditions in connection with prior distributions. We say the prior distribution Γ satisfies condition I if the common boundary D is in the supports of Γ_i, $i = 1, 2$. The prior satisfies condition II if the supports of Γ_i, $i = 1, 2$ contain $B_i \cap M(\rho, D)$ for some $\rho > 0$. Clearly condition I is equivalent to Ω lying in the supports of $\boldsymbol{\Gamma}_i$, $i = 1, 2$, and condition II is equivalent to the supports of $\boldsymbol{\Gamma}_i$ containing $\Theta_i \cap M(\alpha, \Omega)$, for some $\alpha > 0$, where α is related to ρ.

THEOREM 3.1. *Let the family of densities be given as in* (3.3). *For $p = 1$, let Γ satisfy condition* I, *while for $p > 1$, let Γ satisfy condition* II. *Then there exists $m < \infty$ such that the Bayes test satisfies $N_\Gamma \le m$, with probability one $[P_\theta]$, for all $\theta \in \Theta$, m independent of θ.*

PROOF. (Since the proof of the theorem is formidable it may prove helpful to the reader to work with the special cases of multivariate and univariate normality with unknown means and covariance identity.)

We may assume $0 < \pi_1 < 1$. Let $\beta_0 \in$ D, let ϵ satisfy the hypotheses of Lemma 2.2, and let $M_i = M(\epsilon, \beta_0) \cap B_i$. Then

$$\Gamma_i(M_i \mid \bar{X}_n = x) = \int_{M_i} \bar{f}_{\beta n}(x)\Gamma_i(d\beta) \Big/ \left[\int_{M_i} \bar{f}_{\beta n}(x)\Gamma_i(d\beta) + \int_{B_i - M_i} \bar{f}_{\beta n}(x)\Gamma_i(d\beta)\right], \tag{3.9}$$

where x represents a point in the p-dimensional sample space of $\bar{X}_n$. Use (3.3) in (3.9) and rewrite so that (3.9) becomes

$$1 \Big/ \left[1 + \left\{\int_{B_i - M_i} e^{-nI(x,\beta)}\Gamma_i(d\beta) \Big/ \int_{M_i} e^{-nI(x,\beta)}\,\Gamma_i(d\beta)\right\}\right]. \tag{3.10}$$

Since D is contained in the support of Γ_i, the curly bracketed term in (3.10) is bounded above by

$$\{\Gamma_i(M(\delta, \beta_0) \cap B_i)\}^{-1} e^{-n\{\inf I(x,\beta):\beta \in B_i - M_i\}} / e^{-n\{\sup I(x,\beta):\beta \in M(\delta,\beta_0)\}}, \tag{3.11}$$

for any δ, $0 < \delta < \epsilon$. Note for the numerator of (3.11) that for $x \in M(\delta, \beta_0)$

$$\epsilon \le \|\beta - \beta_0\| \le \|\beta - x\| + \|x - \beta_0\| \le \|\beta - x\| + \delta, \tag{3.12}$$

so that $\|\beta - x\| \ge \epsilon - \delta$. For the denominator if $x \in M(\delta, \beta_0)$, $\|x - \beta\| \le 2\delta$. Using (3.8) with $H = \bar{M}(\epsilon, D)$, and noting that δ can be chosen so that $A_5(\epsilon - \delta)^2 > A_6(2\delta)^2$, it follows that for all $x \in M(\delta, \beta_0)$, there exists $n(\beta_0)$ sufficiently large so that if $n \ge n(\beta_0)$, (3.9) > $(1 - \epsilon)$ for $i = 1, 2$ and (2.10) holds. We wish to select n so as not to depend on β_0. Therefore fix γ, $0 < \gamma < \delta$ and consider $\{M(\delta, \beta) : \beta \in D)\}$. This is an open cover of $\bar{M}(\gamma, D)$. This latter set is compact, which implies there exists a finite subcover $M(\delta, \beta_j)$, $j = 1, 2, \cdots, J$. Let $n_1 = \max n(\beta_j)$, and note that for $n \ge n_1$ and all $x \in M(\gamma, D)$, (3.9) > $(1 - \epsilon)$ and (2.10) holds. Consequently, by Lemma 2.2, δ_Γ stops whenever $n \ge n_1$ and $x \in M(\gamma, D)$.

Next we show that an integer m_1 can be chosen so that for all $n \ge m_1$, $\tau_n(x) < c$ if $x \in M(\gamma, D)^{\sim}$ where $^{\sim}$ denotes complement in B. If such an m_1 is found, by choosing $m = \max(n_1, m_1)$ the theorem will be proved. Recognize by the definition of $\tau_n(x) = \min_{1\le i\le 2}(\pi_i g_{in}(x))$, that $\tau_n(x) < c$ if $\pi_2 g_{2n}(x) < c$ or $\pi_2 g_{2n}(x) > 1 - c$. Since $\pi_2 g_{2n}(x)$ can be written as

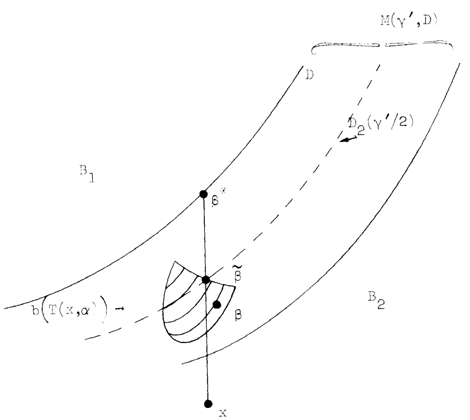

FIG. 1

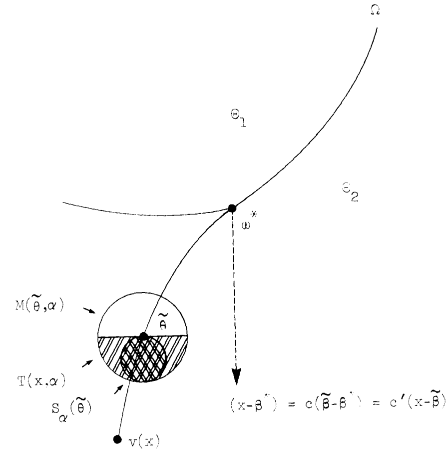

FIG. 2

$$1\Big/\left[1+(\pi_1/\pi_2)\left\{\int_{B_1} e^{-nI(x,\beta)}\,\Gamma_1(d\beta)\Big/\int_{B_2} e^{-nI(x,\beta)}\Gamma_2(d\beta)\right\}\right], \tag{3.13}$$

the theorem will be proved if we show that the curly bracketed expression in (3.13) tends to zero or to ∞ as $n \to \infty$ uniformly for $x \in M(\gamma, D)^{\sim}$. Our plan is as follows: Partition $M(\gamma, D)^{\sim}$ into two sets, namely $B_2 - M(\gamma, D)$ and $B_1 - M(\gamma, D)$. For $x \in B_2 - M(\gamma, D)$, find m_2 sufficiently large so that the curly bracketed expression of (3.13) is arbitrarily close to zero for $n \geq m_2$. For $x \in B_1 - M(\gamma, D)$, find m_3 sufficiently large so that the curly bracketed expression of (3.13) is arbitrarily large for $n \geq m_3$. Then choose $m_1 = \max(m_2, m_3)$. The argument for $x \in B_1 - M(\gamma, D)$ would be completely analagous to the argument for $x \in B_2 - M(\gamma, D)$. Hence we argue only the case where $x \in B_2 - M(\gamma, D)$.

To summarize, the proof will be complete if we can show that the curly bracketed expression of (3.13) tends to zero uniformly for $x \in B_2 - M(\gamma, D)$.

Let Γ satisfy condition II and let $\gamma' = \min(\gamma, \rho)$ where ρ is described in condition II. The set D is compact, hence the set $D_2(\gamma'/2) = \partial M(\gamma'/2, D) \cap B_2$ is compact. (Here, ∂ denotes boundary and so $D_2(\gamma'/2)$ can also be written as $\{\beta : \beta \in B_2, \|\beta - D\| = \gamma'/2\}$.) From continuity and convexity properties of I and compactness of D_2 and D it follows that there exists $\eta > 0$ and $\epsilon > 0$ such that $\beta' \in D_2(\gamma'/2)$, $\beta \in M(\beta', \eta)$, $\beta_1 \in (B_1 \cup D)$, implies

$$I(\beta', \beta_1) - I(\beta', \beta) > \epsilon > 0. \tag{3.14}$$

Obviously, $\eta \leq \gamma'/2$. Since the map $b : \Theta \to B$ is continuous and $D_2(\gamma'/2)$ is compact there exists an $\alpha > 0$ such that $b(\theta') \in D_2(\gamma'/2)$ and $\|\theta - \theta'\| < \alpha$ implies $\|b(\theta) - b(\theta')\| \leq \eta$.

Now, define $\beta^*(x) \in \bar{B}_1$ by

$$I(x, \beta^*) = \inf_{\beta \in B_1} I(x, \beta). \tag{3.15}$$

(When (3.15) does not uniquely define β^*, then any choice for $\beta^*(x)$ satisfying (3.15) is okay.) Recall that Θ is convex, $I^*(v(x), \theta) = I(x, b(\theta))$ is convex in θ, and the expectation map b, is continuous and 1-1. These facts imply that for $x \in B_2$, $\beta^*(x)$ lies in D. For $x \notin M(\gamma', D)$ define $\tilde{\beta}(x)$ to be the point closest to $\beta^*(x)$ which lies in the intersection of $D_2(\gamma'/2)$ and the line joining x to $\beta^*(x)$. Formally, $\tilde{\beta}(x) = \beta_{a_x}(x)$ where $\tilde{\beta}_a(x) = \beta^*(x) + a(x - \beta^*(x))$ and $a_x = \inf\{a : \tilde{\beta}_a(x) \in M(\gamma', D) - M(\gamma'/2, D)\}$. Figure 1 illustrates x, β^*,

$\tilde{\beta}$, etc., and Figure 2 illustrates the corresponding points $v(x)$, $\omega^*(x) = v(\beta^*(x))$, $\tilde{\theta}(x) = v(\tilde{\beta}(x))$, etc. (These figures are drawn for the case where Ω is a differentiable manifold or curve, in which case the (outward) normal to Ω at ω^* is in the direction of $x - \beta^*$. See Efron (1978) page 365).

Let $T(x, \alpha) = \{\theta\colon \theta \in \Theta, \theta \in M(\tilde{\theta}(x), \alpha), (\theta - \tilde{\theta})'(x - \tilde{\beta}) > 0\}$. Again, see Figure 2. Note for later use that $b(T(x, \alpha)) \subset M(\gamma', D)$ since $\eta < \gamma'/2$. Hence $T(x, \alpha) \subset \operatorname{supp} \boldsymbol{\Gamma}_2$. Also note that $T(x, \alpha)$ contains the interior of a ball of radius $\alpha/2$. Call this ball $S_\alpha(\tilde{\theta})$. (We treat the case where D_2 is a connected set. When D_2 is finite the argument throughout is much easier. When D_2 is not connected and not finite the argument just requires repetitions.) Note that $S_\alpha(\tilde{\theta})$ converges pointwise to $S_\alpha(\tilde{\theta}_0)$ if $\| \tilde{\theta} - \tilde{\theta}_0 \| \to 0$. Hence there is a constant $\xi > 0$ such that

$$\boldsymbol{\Gamma}_2(T(x, \alpha)) > \xi > 0, \tag{3.16}$$

for all $x \notin M(\gamma', D)$. To prove (3.16) note that $\inf_{x \notin M(\gamma', D)} \boldsymbol{\Gamma}_2(T(x, \alpha)) \geq \inf_{\tilde{\theta} \in D_2} \boldsymbol{\Gamma}_2(S_\alpha(\tilde{\theta}))$. Express $\boldsymbol{\Gamma}_2(S_\alpha(\tilde{\theta}))$ as the integral of the indicator function of the open set $S_\alpha(\tilde{\theta})$ and use Fatou's lemma to conclude that $\boldsymbol{\Gamma}_2(S_\alpha(\tilde{\theta}))$ is a lower semicontinuous function of $\tilde{\theta}$. The inequality in (3.16) then follows since D_2 is compact and $\boldsymbol{\Gamma}_2(S_\alpha(\tilde{\theta})) > 0$ for all $\tilde{\theta} \in D_2$ by condition II.

We are now ready to proceed to show that the curly bracketed expression of (3.13) tends to zero uniformly for $x \in B_2 - M(\gamma, D)$. Let $\theta \in T(x, \alpha)$. Applying (3.4) yields

$$\begin{aligned} I(x, \beta^*) - I(x, \beta) &= (\theta - \omega^*)'x - (\psi(\theta) - \psi(\omega^*)) \\ &= (\theta - \omega^*)'\tilde{\beta} - (\psi(\theta) - \psi(\omega^*)) + (\theta - \omega^*)'(x - \tilde{\beta}) \\ &= I(\tilde{\beta}, \beta^*) - I(\tilde{\beta}, \beta) + (\theta - \omega^*)'(x - \tilde{\beta}). \end{aligned} \tag{3.17}$$

Now, $(x - \tilde{\beta}) = \| x - \tilde{\beta} \| (\tilde{\beta} - \beta^*)/\| \tilde{\beta} - \beta^* \| = C(\tilde{\beta} - \beta^*)$, where $C \geq 1$. Hence

$$(\theta - \omega^*)'(x - \tilde{\beta}) = (\theta - \tilde{\theta})'(x - \tilde{\beta}) + (\tilde{\theta} - \omega^*)'(x - \tilde{\beta}) \geq C(\tilde{\theta} - \omega^*)'(\tilde{\beta} - \beta^*), \tag{3.18}$$

because of the above and the definition of $T(x, \alpha)$. Next,

$$(\tilde{\theta} - \omega^*)'(\tilde{\beta} - \beta^*) = (\tilde{\theta} - \omega^*)'(b(\tilde{\theta}) - b(\omega^*)) > 0 \tag{3.19}$$

since $b(\theta) = \psi'(\theta)$ and ψ is a positive, convex function. Thus

$$\begin{aligned} I(x, \beta^*) - I(x, \beta) &= I(\tilde{\beta}, \beta^*) - I(\tilde{\beta}, \beta) + (\theta - \omega^*)'(x - \tilde{\beta}) \\ &\geq I(\tilde{\beta}, \beta^*) - I(\tilde{\beta}, \beta) > \epsilon > 0, \end{aligned} \tag{3.20}$$

by (3.14).

Return to the curly bracketed expression in (3.13):

$$\begin{aligned} &\left\{ \int_{B_1} e^{-nI(x,\beta)} \Gamma_1(d\beta) \Big/ \int_{B_2} e^{-nI(x,\beta)} \Gamma_2(d\beta) \right\} \\ &\leq e^{-nI(x,\beta^*)} \Big/ \int_{b(T(x,\alpha))} e^{-nI(x,\beta)} \Gamma_2(d\beta) \\ &= \left(\int_{b(T(x,\alpha))} e^{-n[I(x,\beta) - I(x,\beta^*)]} \Gamma_2(d\beta) \right)^{-1} \\ &\leq \left(\int_{b(T(x,\alpha))} e^{n\epsilon} \Gamma_2(d\beta) \right)^{-1} \\ &= e^{-n\epsilon} (\boldsymbol{\Gamma}_2(T(x, \alpha)))^{-1} \leq \xi^{-1} e^{-n\epsilon} \to 0 \end{aligned} \tag{3.21}$$

uniformly as $n \to \infty$ for all $x \notin M(\gamma', D)$. This is the desired result for $p > 1$.

If $p = 1$, we need consider the case where one of the priors, say Γ_1, puts positive mass on a boundary point, say β', where $\beta' \in B_1$. For simplicity of exposition we treat the one sided hypothesis testing problem, although all the cases can be argued similarly but with more detail. Thus we assume $x \in M(\gamma, \beta')^{\sim}$, $x \in B_1$ and $B_2 = (\beta', \infty)$. In this case the reciprocal of the bracketed expression in (3.13) is bounded above by

$$K \int_{B_2} e^{-nI(x,\beta)}\Gamma_2(d\beta)/e^{-nI(x,\beta')}. \tag{3.22}$$

We may choose ζ sufficiently small so that for any given arbitrarily small φ, (3.22) can be written as

$$\begin{aligned} K\left[\int_{\beta'}^{\beta'+\zeta} e^{-nI(x,\beta)}\Gamma_2(d\beta) + \int_{\beta'+\zeta}^{\infty} e^{-nI(x,\beta)}\Gamma_2(d\beta)/e^{-nI(x,\beta')}\right. \\ \le K\{\Gamma_2(\beta', \beta' + \zeta) + e^{-n[I(x,\beta'+\zeta)-I(x,\beta')]}\} \\ \le K\{\varphi + e^{-n[I(x,\beta'+\zeta)-I(x,\beta')]}\}. \end{aligned} \tag{3.23}$$

By properly choosing ζ and n, using the properties of I as a function of β for fixed x and as a function of x, the expression on the right-hand side of (3.23) can be chosen as close to zero as desired, uniformly in x, for $x \in B_1$. The remainder of the argument is as in the case $p > 1$. □

REMARK 3.1. The requirement that Θ is convex can be replaced by the following weaker condition: If $x \in M(\gamma, D)^{\sim}$ then the line joining $v(x)$ and ω^* must line in Θ.

REMARK 3.2. When $p = 1$ and $\Theta_1 \cap \Theta_2 \neq \emptyset$ condition II and the other conditions of Theorem 3.1 imply bounded stopping time for the Bayes test. This case is of interest since in Brown, Cohen, and Strawderman (1980) a complete class characterization for testing Θ_1 vs Θ_2 is given in terms of Bayes procedures or procedures closely related to Bayes procedures for testing $\bar{\Theta}_1$ vs $\bar{\Theta}_2$.

REMARK 3.3. For testing the noncentrality parameter of a noncentral chi-square distribution condition II again yields that the Bayes test has bounded stopping time. This follows from the fact that the theorem works for the multivariate normal distribution with unknown mean and known covariance matrix, which without loss of generality is chosen to be the identity matrix. Spherically symmetric priors yield tests based on a chi-square random variable so that spherically symmetric priors satisfying condition II yield the desired result.

We conclude this section with discussion, extensions, and generalizations. If $p = 1$, Theorem 3.1 can be proved for a class of distributions more general than the exponential family: Observations need be independent and identically distributed and there must be a sequence $\{T_n\}$ of one dimensional sufficient statistics whose distribution has monotone likelihood ratio. Such a class includes the exponential family as well as the uniform distribution with unknown scale parameter. The proof for this class of distributions is a probabalistic proof that is completely different than the proof of Theorem 3.1. Since the case of the uniform distribution with unknown scale parameter can be verified by directly referring to the fundamental lemmas and since there are not many other interesting distributions in this class we will not give a detailed proof. The proof does, of course, refer to the fundamental lemmas and the bracketed expressions in (3.10) and (3.13) are then studied. Both of these expressions, regarded as random variables, are martingales. The martingale convergence theorem and monotonicity properties of the ratios are then used to show that (3.10) and (3.13) have the desired limiting values.

Theorem 3.1 is quite general in the sense that it immediately identifies a large class of prior distributions for which the Bayes tests has a bounded stopping time. The result applies to the multivariate exponential family provided other conditions on the null and alternative spaces are satisfied. Discrete and continuous distributions are permitted.

The result can be obtained for more general loss functions. Let $W(\theta, \tau)$ represent the loss due to terminal decision. Then with our present loss function $W(\beta, \tau) = 0$ whenever $(\beta \in B_1, \tau = 1)$ and $(\beta \in B_2, \tau = 2)$; $W(\beta, \tau) = 1$ if $(\beta \in B_1, \tau = 2)$ and $(\beta \in B_2, \tau = 1)$. First we give

COROLLARY 3.1. *Assume all the conditions of Lemmas* 2.1, 2.2, *and Theorem* 3.1 *are true. Let the loss function be as before except that now* $W(\beta, \tau) = w_1 > 0$ *if* $(\beta \in B_1, \tau = 2)$ *and* $W(\beta, \tau) = w_2 > 0$ *if* $(\beta \in B_2, \tau = 1)$. *Then the conclusion of Theorem* 3.1 *is true.*

PROOF. Without loss of generality assume $w_1 \le w_2$. In Lemma 2.2 choose $k \ge w_1/c$, $0 < \xi < c/w_1(k+1)$, and $0 < \epsilon < (c/w_1) - (k+1)\xi$. The proof of Lemmas 2.1, 2.2, and Theorem 3.1 can be repeated. □

Now let $p = 1$ and consider the one sided testing problem. That is, $B_1 = (-\infty, \beta^*]$, $B_2 = (\beta^*, \infty)$. Let $W(\beta, \tau) = 0$ if $(\beta \in B_1, \tau = 1)$ or $(\beta \in B_2, \tau = 2)$. If $(\beta \in B_1, \tau = 2)$, $W(\beta, \tau) = w_1(\beta) \ge 0$ while if $(\beta \in B_2, \tau = 1)$, $W(\beta, \tau) = w_2(\beta) \ge 0$. Assume for $i = 1, 2$

$$w_i(\beta) \text{ continuous}, \quad w_i(\beta^*) = 0, \qquad w_i(\beta) > 0 \quad \text{except at} \quad \beta = \beta^*. \tag{3.24}$$

Consider

$$\int_{B_i} w_i(\beta)\bar{f}_{\beta n}(x)\Gamma_i(d\beta) \Big/ \left[\pi_1 \int_{B_1} \bar{f}_{\beta n}(x)\Gamma_1(d\beta) + \pi_2 \int_{B_2} \bar{f}_{\beta n}(x)\Gamma_2(d\beta)\right], \tag{3.25}$$

$$\int_{\beta^*+\epsilon}^{\infty} w_2(\beta)\bar{f}_{\beta n}(x)\Gamma_2(d\beta) \Big/ \int_{\beta^*}^{\beta^*+\epsilon} \bar{f}_{\beta n}(x)\Gamma_2(d\beta), \tag{3.26}$$

and

$$\int_{-\infty}^{\beta^*-\epsilon} w_1(\beta)\bar{f}_{\beta n}(x)\Gamma_1(d\beta) \Big/ \int_{\beta^*-\epsilon}^{\beta^*} \bar{f}_{\beta n}(x)\Gamma(d\beta), \tag{3.27}$$

where ϵ is any positive number.

COROLLARY 3.2. *Assume the model for Theorem* 3.1 *holds for* $p = 1$ *except now the loss function satisfies* (3.24). *Assume* Γ *satisfies condition* II. *Assume* (3.25) *is finite for all* x *and* n. *Furthermore assume that for any given* $\epsilon > 0$ *there exists a positive number* γ *such that for all* x *such that* $x \in (\beta^*, \beta^* + \gamma)$ $\lim_{n\to\infty}$ (3.26) $= 0$ *and for all* $x \in (\beta^*, \beta^* - \gamma)$ $\lim_{n\to\infty}$ (3.27) $= 0$. *Then the Bayes test has bounded stopping time.*

REMARK 3.4. The conditions in the statement of the corollary pertaining to (3.25), (3.26), and (3.27) are very easy to check and are not very restrictive. They hold when $w_i(\beta)$ are bounded; when $\int_{B_i} w_i(\beta)\Gamma_i(d\beta) < \infty$; and in many cases when neither of these two more restrictive conditions hold. The verification of (3.26) and (3.27) would involve an argument similar to that used in the beginning of the proof of Theorem 3.1. (see the curly bracketed expression in (3.10)). The $w_i(\beta)$ would have to tend to infinity very fast if conditions (3.26) and (3.27) are to be violated. Note that a linear loss satisfies (3.24).

PROOF OF COROLLARY 3.2. The proof involves showing that for all x, $\tau_n(x) < c$ for all n sufficiently large. This is accomplished, first by showing that $\tau_n(x)$ tends to zero uniformly for $x \in (\beta^* + \gamma, \infty)$ and uniformly for $x \in (-\infty, \beta^* - \gamma)$. These two facts are proven as they were in Theorem 3.1 except that now (3.24) and condition II are needed. For $x \in (\beta^* - \gamma, \beta^* + \gamma)$ we argue as follows:

By definition

$$\tau_n(x) = \min_{1\le i\le 2} \left\{ \int_{B_i} w_i(\beta)\bar{f}_{\beta n}(x)\Gamma_i(d\beta) \Big/ \left[\pi_1 \int_{B_1} \bar{f}_{\beta n}(x)\Gamma_1(d\beta) + \pi_2 \int_{B_2} \bar{f}_{\beta n}(x)\Gamma_2(d\beta) \right] \right\}. \tag{3.28}$$

(Note that $\tau_n(x)$ is now defined differently from how it was before.) Split up the range of integration for the numerator in (3.28) into two pieces $((\beta^*, \beta^* + \epsilon), (\beta^* + \epsilon, \infty))$ for B_2), use conditions (3.24), (3.25), and (3.26) and the result regarding $\tau_n(x) < c$ for n sufficiently large follows. □

Next replace (3.24) by

$$\begin{aligned} &w_i(\beta) \text{ continuous, } \lim_{\beta\to\beta^*} w_i(\beta) = w > 0 \qquad w_i(\beta) > A > 0, \\ &\sup_{\beta\in B_i} w_i(\beta)\bar{f}_{\beta n}(x) < \infty \qquad \text{for all } x. \end{aligned} \tag{3.29}$$

COROLLARY 3.3. *Assume the model for Theorem* 2.1 *holds for* $p = 1$ *except now the loss function satisfies* (3.29). *Assume* (3.25) *is bounded for all x and n. Then the Bayes test has bounded stopping time.*

PROOF. The analogue of Lemma 2.1 is needed and can be produced with condition (3.29). The analogue of Lemma 2.2 is needed and can be produced when (3.25) is bounded for all x and n. The arguments of Theorem 3.1 can then be used with minor alterations. □

COROLLARY 3.4. *Assume the model of Theorem* 3.1 *except now let the cost of the* jth *observation be* $c_j > 0$. *Assume* $c_j > c > 0$. *Then the conclusion of Theorem* 3.1 *holds.*

PROOF. The only changes required are in Lemma 2.2 where k, ξ, and ϵ are now chosen so that $k \ge \frac{1}{2} c$, $0 < \xi < 2c/(k + 1)$ and $0 < \epsilon < 2c - (k + 1)\xi$. □

REFERENCES

BERK, R. H., BROWN, L. D. and COHEN, A. (1981). Properties of Bayes Sequential Tests. *Ann. Statist.* **9** 678–682.

BROWN, L. D. and COHEN, A. (1981). Inadmissibility of large classes of sequential tests. *Ann. Statist.* **9**. To appear.

BROWN, L. D., COHEN, ARTHUR. and STRAWDERMAN, W. E. (1980). Complete classes for sequential tests of hypotheses. *Ann. Statist.* **8** 377–398.

EFRON, B. (1978). The geometry of exponential families. *Ann. Statist.* **6** 362–376.

KALLENBERG, W. C. M. (1978). *Asymptotic Optimality of Likelihood Ratio Tests in Exponential Families.* Mathematical Centre Tracts 77, Amsterdam.

LE CAM, L. (1955). An extension of Wald's theory of statistical decision functions. *Ann. Math. Statist.* **26** 69–81.

RAY, S. N. (1965). Bounds on the maximum sample size of a Bayes sequential procedure. *Ann. Math. Statist.* **36** 859–878.

ROCKAFELLAR, R. T. (1970). *Convex Analysis.* Princeton Univ.

SCHWARZ, G. (1962). Asymptotic shapes of Bayes sequential testing regions. *Ann. Math. Statist.* **33** 224–236.

SOBEL, M. (1952). An essentially complete class of decision functions for certain standard sequential problems. *Ann. Math. Statist.* **23** 319–337.

R. H. BERK
ARTHUR COHEN
DEPARTMENT OF STATISTICS
RUTGERS UNVERSITY
NEW BRUNSWICK, NEW JERSEY 08903

L. D. BROWN
MATHEMATICS DEPARTMENT
CORNELL UNIVERSITY
ITHACA, N.Y. 14850